

Final White Paper

Digital Editions, Digital Corpora and New Possibilities for the Humanities in the Academy and Beyond

HT-256958-17

12/30/2019

Project Activities

For the two-year duration of this project, the major activities of the Digital Editions team involved planning and hosting both a two-week NEH Institute in the summer of 2018, and a follow up digital editions workshop held in the spring of 2019.

In December of 2017, the summer institute was announced on the Perseus Digital Library¹ website, various social media sites and sent out to topically related listservs. In mid-January 2018, a second request for applications was sent out with a final deadline of February 1st, 2018. The announcement noted that the institute was centered on advanced topics and particularly encouraged applications from participants interested in integrating contributions from undergraduates and expanding the global presence of the humanities. A website² was created for the institute and included the full text of the CFP and application instructions. Applying for the institute required completing a registration form and submitting both a CV and a two page statement of interest.

Over 55 registrations were initially received with 48 full submissions received by the deadline. Applications were received from around the world and included undergraduate and graduate students, postdocs, junior and senior faculty, librarians, and several independent scholars. Classical scholars did make up a significant number of the applications but other disciplinary fields were also well represented. These included history, English language and literature, medieval studies, Islamic studies, library science, education, and criminology. In addition, the languages for study were also quite diverse with Ancient Greek and Latin having the largest number of applications, but there was also a strong concentration of scholars working in modern and middle English, Arabic, Persian, Hebrew, Portuguese, and Sanskrit.

The institute organizers ultimately selected a group of applicants whom they felt shared topical interests and languages of study that would both complement one another and meet the stated goals of the institute. Two determining criteria that were used in the final selection were a willingness to make one's work openly available and a desire to support undergraduate work in the digital humanities. Notifications of acceptance were sent out in March 2018, and in April 2018 there were 29 confirmed participants (with 15 based outside of the United States). In the last weeks before the institute, however, five participants (one local, three international, and one out of state visiting scholar) withdrew from the institute. Three of the five informed the organizers a week prior

¹ <https://www.perseus.tufts.edu/>

² <https://sites.tufts.edu/digitaleditions/>

to the start of the institute making it impractical to offer their spaces to anyone else. There were thus only 24 final institute participants resulting in a smaller group than originally anticipated. The final group that attended in July represented four countries other than the United States: Brazil, Croatia, France, Germany. The disciplinary range also varied and included classics, medieval studies, Middle Eastern studies, educational technology, history, and Slavic languages. A number of different historical languages were represented, such as Arabic, Ancient Greek, Croatian, Latin, Persian, and Sanskrit. Participants included an undergraduate student, a number of doctoral students, research associates, librarians, full and junior professors, and the director of a digital humanities center. A total of 32 people participated in the institute, including attendees, organizers, guests, and staff.

The institute, “Digital Editions, Digital Corpora and New Possibilities for the Humanities in the Academy and Beyond”, was held from July 16 to July 27, 2018. The institute website was used to make major announcements during the two weeks and a Moodle³ course was developed for the institute and hosted at Humboldt University-Berlin. This course⁴ was used throughout the duration of the institute to host content (including daily class slides, tool links, and participant presentations), communicate with participants, and send out updates as needed.

On the first day, participants presented overviews of their research questions (if they had not decided on one, they presented on what they hoped to learn) to help the group to get to know each other and to aid instructors in discovering common participant interests and the varying levels of technical skill the attendees possessed. During the institute, the participants received extensive instruction in how to create annotated digital corpora and digital editions. The first week involved a mix of teaching and hands on practice with TEI-XML encoding as well as a number of corpus annotation tools and technologies (ANNIS⁵, Oxygen XML editor⁶, EXMARaLDA annotation editor).⁷ As needed, instructors assisted participants in convert data for use with ANNIS and gave the group a specific text to enter as a learning exercise at the end of the first week. Participant work was done both individually and in collaborative groups. During the second week, instruction continued in the creation of digital editions and corpora. Additionally participants were also introduced to a number of important scholarly research questions that can be explored through the use of digital corpora including text reuse and parallel text/translation analysis. Instruction then concluded with an overview of the importance of open access data publication, data preservation repositories, licensing issues, and how to teach with the editions/corpora that have been created. The final day of the institute was then dedicated to student presentations.

³ <https://moodle.hu-berlin.de/course/view.php?id=81147>

⁴ The entire course site can be viewed as an anonymous guest and the majority of the learning content (other than the private user forum) may be viewed or downloaded.

⁵ ANNIS (<http://corpus-tools.org/annis/>) is an “open source, cross platform (Linux, Mac, Windows), web browser-based search and visualization architecture for complex multi-layer linguistic corpora with diverse types of annotation”

⁶ <https://www.oxygenxml.com/>

⁷ <http://exmaralda.org/en/partitur-editor-en/>

After the conclusion of the 2018 institute, project work next turned to planning the follow up workshop in the spring of 2019. Several meetings between project staff and the PIs were held to discuss different approaches for the follow up workshop regarding format, topic and length of time. Ultimately a two day workshop was designed, entitled “Digital Editions in Practice,” and much of the work in late winter 2018 to early spring 2019 involved inviting instructors to teach on relevant topics. After confirming relevant instructors for the desired topics, the project team also contacted previous institute attendees to inform them about the spring workshop and to invite their collaboration and or participation. Only one previous attendee indicated an interest in and ultimately attended the 2019 workshop

The workshop CFP was sent on January 31, 2019 to various disciplinary listservs and DH mailing groups, as well as being posted on DH Now. Potential attendees were asked to fill out an online application that asked them to indicate topics of interest, technology experience, the materials with which they wished to work and goals for the workshop. A total of 36 applications were received for the workshop and one applicant requested adding a colleague, making for 37 responses to the call. While 33 participants were initially accepted (ones that were not accepted were due to insufficient project information or incompatible subject areas of work and goals), by the time of the workshop five applicants had withdrawn or not responded to their acceptance, and another wanted to participate remotely. There were also several last minute requests (and walk-ons) from local applicants that were accepted, so in total there were 33 official attendees at the 2019 workshop. The final group of participants, as with the institute before it, included librarians and library school students, graduate students in various disciplines, postdoctoral research associates in the digital humanities, faculty from several disciplines (including retired faculty and an independent scholar), the director of a digital humanities institute, and individuals involved in commercial digital scholarly publishing.

The “Digital Editions in Practice” workshop was held from May 31-June 1, 2019. The first day involved hand on presentations of different digital technologies used in the creation of digital editions including (International Image Interoperability Framework,⁸ OCR, CTS/CITE architecture⁹, INCEpTION,¹⁰ named entity recognition, Recogito¹¹, topic modeling), and the second day featured small breakout groups where participants decided what topic they wished to pursue in further detail. After the workshop, final project activities involved managing travel reimbursements and other administrative details.

Overall, there were no major changes in project activities during the performance period and there were also no major changes in key personnel, other than that Anke Luedeling was not able to

⁸ <https://iiif.io/>

⁹ <http://cite-architecture.org/>

¹⁰ <https://inception-project.github.io/>

¹¹ <https://recogito.pelagios.org/>

participate in the spring 2019 workshop due to teaching responsibilities and scheduling conflicts in Germany.

Accomplishments

The original grant application proposed to provide an opportunity for 30 humanists to spend two weeks at an intensive institute studying a range of digital methods for annotating textual sources, and this goal was largely met, although a smaller number ended up attending than desired due to last minute withdrawals and receiving fewer applications than expected. The proposal also outlined that a follow up workshop would be held in 2019, and this goal was accomplished as well.

Additionally the application proposed that the institute curriculum would be both informed by and influence the Sunoikisis DC¹² program, an international consortium of digital classics programs that offers collaborative online courses that allow students to pursue interdisciplinary learning at the intersection of humanities and computer science. This goal was also achieved as the institute exposed the participants to a number of SunoikisisDC topics and encouraged the group to work together collaboratively to solve common problems. Similarly, the original proposal stated that the institute aimed “at focusing on two complementary topics corresponding roughly to “distant reading” and “close reading” and wanted to help participants understand the importance of both exploring automated methods for large scale corpora while not losing sight of the continuing importance of manual scholarly annotation. In terms of this goal, the institute was arguably quite successful, in that participants received both rigorous instruction in terms of corpus linguistics and methods of automatic semantic annotation but also spent a great deal of time manually marking up their own texts both individually and in groups as well as exploring the use of automated technologies against those texts.

The original proposal also noted that the institute directors would encourage applications from those working with a wide variety of linguistic sources and from a variety of professional backgrounds, a goal that was also largely realized. The range of languages studied included not just the classical languages of Greek and Latin, but also Sanskrit, Arabic, Persian Hebrew, and several modern languages including French, English and Portuguese.

In the original application, five questions were listed that would guide technological instruction and discussion at the institute: How do digital technologies change research methods, questions and results of each participant? To what extent can our research advance, and draw upon, intellectual life beyond the academy and citizen scholarship? How do these technologies affect the formal teaching that participants can offer? What are the implications of conducting research and teaching within a global network? Are there new opportunities whereby those who work primarily in English can make their ideas more accessible beyond the English speaking world? All of these questions were explored at length during the institute, as participants were encouraged to find their own answers in terms of their own work and research goals. In the final presentations given

¹² <http://www.dh.uni-leipzig.de/wo/sunoikisisdc/>

by each participant on the last day of the institute, numerous attendees commented on how they had found or mastered a new method with which to pursue their digital scholarship, spoke about their renewed commitment to publishing their research as open access so that it would reach a larger audience, and demonstrated a greater openness in terms of finding and engaging with new research collaborators, including in some cases, their students.

Audiences

The intended audience for this project was always international in scope and had no specific focus on age, race, or gender. Although the announcement did explicitly state that the institute and workshop were aimed at humanists, the project team encouraged applications not just from faculty but also from all levels of students, and from library and museum professionals. While the project was directed at applicants with at least some level of digital humanities experience, the only essential requirements were a willingness to experiment with digital technologies in one's research and teaching, a desire to publish one's research openly, and a belief in the need to expand access to and participation in the humanities.

Applicants for both the institute and follow up workshop included as discussed briefly above undergraduate and graduate students, postdocs, junior and senior faculty, library professionals, and independent scholars, all of whose work spanned the humanities disciplines. For the 2018 institute, there were 55 initial applications, with 48 final and complete submissions by the deadline. While 29 applicants were accepted, due to a number of withdrawals, there were 24 full-time participants at the institute. For the 2019 workshop, there were 36 formal applications (with one add-on for a colleague), as with the institute there were a number of withdrawals, some last minute local additions, and a final group of 33 full participants. A total of 68 individuals, including project PIs, staff, and presenters participated throughout the course of this grant (a full list of which can be found in the appendices).

Participants included a number of faculty from different disciplines, not only classical studies, but also including history, English, French, social sciences and computer science. There were a large number of PhD students from classical studies but also from middle and near eastern studies, religion, corpus linguistics, and library science. A number of full time library professionals also attended with diverse disciplinary backgrounds including not just classical studies, but medieval history, Islamic studies, and educational technology. In addition to these attendees, there were several independent scholars, several postdocs in both classics and the digital humanities, a director of a digital humanities center, and one particularly enterprising and impressive undergraduate student in classics. The historical languages of study were quite diverse with a strong concentration in Ancient Greek and Latin, but also including Akkadian, Arabic, medieval French, middle English, Hebrew, Sanskrit, Persian and Sumerian. Modern languages of study included Croatian, Dutch, English, French, German, Italian, Portuguese, Turkish, and Russian.

Participants for both the workshop and the institute also came from a reasonably diverse geographic area, with a large number of participants from Europe (Croatia, England, France, Germany, Ireland, Italy and the Netherlands), one from Canada, two from Brazil, two from India, and one from Australia. The American attendees also came from a fairly diverse geographic area (Connecticut, Illinois, Maine, Maryland, Michigan, Missouri, Nebraska, New Jersey, New York, Ohio, Oklahoma, Pennsylvania, South Carolina, Texas, Virginia) with a reasonable number of local attendees from the greater Boston area near Tufts University, host of the events.

Evaluation

Both the 2018 institute and, in particular, the 2019 workshop presented a number of organizational and planning challenges that were not originally anticipated. Hosting the 2018 institute itself presented a number of logistical hurdles. The summer weather was uncooperative and due to a power outage and excessive heat in the room first assigned to the group, a second room had to be secured by the end of the first week. Having to move the participants during the middle of the institute was somewhat daunting and inconvenient, but this did not seem to have an adverse impact overall on the learning experience. Additionally, the follow up workshop held in 2019 faced a number of planning problems, first in finding and then securing commitments by invited lecturers. The workshop also presented some difficulties in implementation, when the two day format came up somewhat short in meeting participant expectations, as will be discussed further below.

Perhaps the most unanticipated challenge, however, was that both the 2018 institute and the 2019 workshop in particular received far fewer applications than were expected. For both the institute and the workshop the organizers attempted to give significant lead time for the application process to encourage as many applications as possible, and also advertised widely on Perseus social media, disciplinary listservs and websites, and posted on sites such as DH Now. Extending the deadline also did not lead to an increased number of applications in either case.

In the original award application, it was noted that for the 2012 institute over 81 applications were received, and it was anticipated that there would be even more than that for this second institute. Instead, the team received only 55 applications, and even though there were a significant number of strong applications, it proved reasonably challenging to select a group that seemed intellectually coherent and suited for advanced instruction. Conversely, as the workshop has such a small number of applicants, the organizers ended up largely accepting everyone who seemed seriously interested in attending, and this led to a (perhaps) overly diverse and intellectually diffuse group that never came together as well as planned.

While the project was not formally or externally evaluated, as outlined in the original grant narrative, participants were asked to evaluate their experience through the use of surveys authored in-house and implemented via Google forms. Two separate participant evaluations were conducted during this project. After the conclusion of both the 2018 institute and the 2019 workshop, participants were asked to complete an evaluation form regarding their experience.

The evaluation form for the 2018 institute allowed for anonymous feedback and asked participants to assess their overall experience, the tools that were used, topics they wish had been included, and how they did or did not plan to use the digital humanities techniques covered in their research, teaching and humanities outreach. After sending a follow up request for evaluation, 8 total responses were received, five participants chose to identify themselves and 3 remained anonymous. Overall the participant assessments were quite positive. The majority of respondents (six of the eight) thought two weeks was the right amount of time; and the thing that most participants liked best about the institute was the chance to meet and work with a diverse group of individuals with interesting digital research projects. While participants opinion of the Moodle course and the ANNIS annotation tool was highly positive, they were less satisfied with Oxygen and suggested having the option of more than one XML editor. On a positive note, all those surveyed planned to implement the tools they learned how to use in their research, with some more limited plans to implement it in their teaching and in digital humanities outreach (largely due to lack of funding, teaching responsibilities in their current position, etc.).

The two most significant criticisms from participants were that too much time was given to student presentations during the course of the institute, and that participants should have been told to install specific software tools such as Oxygen XML ahead of the institute so less time would have been lost on the first day or two to technical issues. These criticisms influenced the organizers preparation for the 2019 workshop. First, it was decided to spend far less time introducing presenters and to only briefly have participants introduce themselves. Also, the application form was modified in order to better gauge how much technical experience participants had before the workshop.

As with the 2018 institute, an evaluation form was sent to workshop participants with the ability to answer anonymously. Thirteen responses were received, with seven participants remaining anonymous. Overall, evaluations for the workshop were more mixed than for the institute. One reason for this was that the format of the workshop on the second day,—initially intended to be guided hands-on learning in small groups,—ended up being more unstructured and less engaging than many participants thought was useful. Several instructors also had to leave unexpectedly early the second day which also complicated the planned format for the second day. Major criticisms received from a number of participants involved a number of recurring themes, including 1) that the technologies being taught were too specialized (particularly the CTS/CITE standard), 2) that the workshop ended up being too focused on the classical disciplines, 3) that there should have been more generic and introductory discussion regarding technologies and workflows used in the creation of digital editions, and 4) that there was not enough of a connection between the presenters and audience, without enough time given to introductions, ice-breaking, and communication between attendees.

From the perspective of the project team, these criticisms of the 2019 workshop seem fair on balance. The 2018 institute arguably was more successful at meeting its goals because of the time available not just for detailed instruction on difficult topics but also for participants to become more familiar with each other and to have the experience of working together—an important component

of much digital humanities work. The organizers' goals for the 2019 workshop may have been either overly ambitious or perhaps not well-articulated in that many participants expressed a feeling that the workshop they attended was not what had been advertised, a crucial point. This key critique serves as a reminder that the project team must better communicate a workshop's end goals to not only the potential attendees but also to the speakers and/or instructors. It has been suggested within the team that the organizers should have provided a somewhat more structured format for speakers and employed better time-management in order to permit more interaction with the participants.¹³ In some areas, it was not clear what the applicability of tools and projects was — and there was not enough designated question and answer time to convey this to the participants. Given the limitations of the two-day workshop format and such a diverse audience, a more cohesive program may have been required. Alternatively, the workshop might have better served this broader group of attendees if it were one or two days longer.

Continuation of the Project

Despite the limited nature of this award, the Perseus Digital Library team and project collaborators continue essential work on several fronts: 1) promoting the innovative use of digital technologies in both the classics and broader humanities, 2) challenging humanists to make their digital work open access, and 3) encouraging humanists (digital and otherwise) to engage more fully with both the broader public and their students. Similarly, the Sunoikisis digital classics program, which particularly inspired the 2018 institute curriculum, continues to thrive at Leipzig University and is about to conclude its 2019 semester with continued planning for the future.¹⁴

In addition, several new important personal collaborations formed during this project. One 2018 workshop attendee is now a full time doctoral student of computer science at Tufts University, and is also actively collaborating with PI Crane on two separate grants. The first, a Tufts Innovates 2019-2020 seed grant,¹⁵ "Working With Corpora: Smart Texts and the Transformation of Reading," has led to the design of a new course at Tufts that simultaneously introduces students in the humanities to advanced computational methods for working with textual corpora and students in computer science to innovative applications of computer science methods in the humanities. This course, a first of its kind at Tufts, directly continues the type of instruction and exploration of the ideas promoted at the institute. The second, a Level 3 NEH Digital Humanities Advancement Grant, "Beyond Translation: New Possibilities for Reading in a Digital Age," began in October 2019, also

¹³ For instance, instead of allowing presenters large blocks of presentation time, the organizers may have better served the audience by 1) restricting the presentations to shorter blocks, 2) asking all presenters to answer key thematic questions (such as "how does this work or tool serve those outside of my field" or "how do we involve undergraduates in this work", etc.) and 3) setting aside more time for presenters to respond to particular use cases.

¹⁴ <https://github.com/SunoikisisDC/SunoikisisDC-2019-2020/wiki/Fall-2019-Sunoikisis-DCH>

¹⁵ <https://provost.tufts.edu/teaching-research/seed-grants-for-teaching-and-research/tufts-innovates/>: "The Tufts Innovates Seed Grant Program seeks to spark imaginative ideas to enhance learning and teaching, including approaches that integrate research and are interdisciplinary. Supported projects are expected to catalyze sustainable initiatives across schools and programs and may lead to further opportunities for internal and external funding."

continues to explore new ways to make ancient language materials more widely accessible to a broader audience through the use of advanced reading technologies and translation tools.

Additionally, an attendee of the 2019 workshop, who also works at Brill publishing as a project manager, is continuing to work with the Perseus Digital Library and its partner project, Open Greek and Latin,¹⁶ to explore how Brill can both make more of its digital content open access and how to make use of some of the digital technologies presented at the workshop in its own content creation stream.

Long Term Impact

The original proposal suggested that the impact of the program would be measured by at least four items: 1) increasing the number of humanists who can integrate methods from computational linguistics in their analysis of textual sources; 2) the production of new data by humanists that can support corpus linguistics; 3) producing more humanists who understand how to integrate their digital research with their teaching; 4) disseminating the results through the Sunoikisis DC curriculum.

The first item of the desiderata was largely accomplished in that there were approximately 57 participants between the institute and the workshop, all of whom were introduced to a number of digital methods. While only a year out from the institute, it will remain to be seen how these seeds take root in the long term careers of participants, particularly in terms of what new digital humanities data may be created (the second of the desiderata) or how they integrate their methods in their teaching where relevant (the third of the desiderata), it can still be reasonably argued that such impact can only be measured in the long-term.

As noted above, the Sunoikisis program continues to flourish, and while it was hoped that some institute participants would contribute to the creation of the Sunoikisis syllabus and become SunoikisisDC instructors, passing on institute lessons to a new audience, so far this has not come to fruition.

One other direct long term impact of this program, as noted above, is that the Perseus Digital Library, has received two grants, one from Tufts, and one from the NEH, to continue exploring the use of corpus linguistics techniques and digital reading technologies, both in the undergraduate classroom and in expanding the reading environment of Perseus itself as well as its new related reading environment, the Scaife Viewer.¹⁷

¹⁶ <https://www.opengreekandlatin.org/>

¹⁷ <https://scaife.perseus.org/>

Award Products

List of websites produced for the grant

Main Institute website: <https://sites.tufts.edu/digitaleditions/>

Moodle course site: <https://moodle.hu-berlin.de/course/view.php?id=81147>

Google Drive for presentations in the 2019 workshop:

https://drive.google.com/drive/folders/1gkffsA6850LFY9_6AUoTaRxQcFVNvLzi